

УДК 004.627

The Compression Algorithm of Hyperspectral Aerospace Images with Use of Mathematical Processing and Intrabands Correlation

Assiya Zh. Sarinova and Alexander V. Zamyatin*

*National Research Tomsk State University
36 Lenin, Tomsk, 634050, Russia*

Received 15.12.2017, received in revised form 29.07.2018, accepted 16.10.2018

This article reveals the features of the channels of hyperspectral aerospace images. There has been suggested the algorithm with interchannel finding of the best channel of intrabands correlation and the use of indexed lossless encoding, which allows to reduce the dimensions of the image channels and to convert them before compression. Also the article shows the results of experiments in comparison with universal archivers such as Winrar, 7Z, and JPEG Lossless according to the compression degree and several parameters: the interchannel correlation, geometric image size, number of channels, number of channel groups. Furthermore, the author of the article studies the options of compression with the search of the main channel of the correlation and without it; these options demonstrate the increase in compression degree due to differential transformations, which allow to store the information in a more compressed form.

Keywords: hyperspectral aerospace images, the compression algorithm, intraband correlation, indexed encoding, channel groups.

Citation: Sarinova A.Zh., Zamyatin A.V. The compression algorithm of hyperspectral aerospace images with use of mathematical processing and intrabands correlation, J. Sib. Fed. Univ. Eng. technol., 2018, 11(8), 882-891. DOI: 10.17516/1999-494X-0110.

© Siberian Federal University. All rights reserved

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

* Corresponding author E-mail address: assiya_prog@mail.ru, zamyatin@mail.tsu.ru

Алгоритм сжатия гиперспектральных аэрокосмических изображений с использованием математической обработки и учетом междиапазонной корреляции

А.Ж. Саринова, А.В. Замятин
Национальный исследовательский
Томский государственный университет
Россия, 634050, Томск, пр. Ленина, 36

В данной статье исследованы особенности каналов гиперспектрального аэрокосмического изображения. Предложен алгоритм с межканальным нахождением наилучшего канала междиапазонной корреляции и применением индексированного кодирования без потерь, позволяющий уменьшить размеры каналов изображения и преобразовать их перед сжатием. Приведены результаты экспериментов в сравнении с универсальными архиваторами Winrar, 7Z и JPEG Lossless по степени сжатия и нескольким параметрам: учету межканальной корреляции, геометрическому размеру изображения, количеству каналов, количеству групп каналов. Рассмотрены варианты сжатия с поиском главного канала корреляции и без него, демонстрирующие повышение степени сжатия за счет разностных преобразований, позволяющие хранить информацию в более уменьшенном виде.

Ключевые слова: гиперспектральные аэрокосмические изображения, алгоритм сжатия, междиапазонная корреляция, индексированное кодирование, группы каналов.

Introduction

Modern satellite-based centers of space monitoring and remote sensing of the Earth (RSE) efficiently accept, register, process, archive and distribute large amounts of data, which are sometimes of hundreds of gigabytes.

It is known that the aerospace images (AI), RSE are characterized by different features: spectral, radiometric, spatial resolutions, geometric sizes of the board, which have different number of channels and ranges.

In the field of compression of hyperspectral AI we already have interesting research results [1-13], however they have some disadvantages. For example, in the works [2, 3] there were presented the approaches to lossless compression of hyperspectral AI using an adaptive prediction and back-search schemes. The proposed scheme is based on the approach of the table prediction and uses two new approaches to improve the compression performance. The first approach uses the spatial correlation of the data and formulates a conclusion of spectral prediction in the filtering process of Vinner. In the second approach, the search scheme is used instead of reference tables, that significantly reduces the storage requirements. The search is greatly reduced using the quantification index. The disadvantage of these approaches is a partial loss of information due to quantification phase.

Some researchers have considered the characteristic features of hyperspectral AI. So, the technique of lossless compression using the spatial-spectral encoding is researched in the following works [1, 4, 7]. The researchers have proposed a new hybrid context prediction for the lossless compression of hyperspectral data, also the spectral correlation matrix is calculated. The authors suggested a new

technology of finding a high correlation of channels groups of hyperspectral data cube on the basis of the group of squared correlation. The disadvantage of such approaches is the time required to compress and restore the image, which depends on calculation of spectral correlation matrix, and also the proposed schemes produce an average compression ratio which reaches only 3,2.

In the article [5, 9] it is proposed to use randomized methods for reducing dimensionality for effective coverage of the correlation structure and encoding the part remained, and for ensuring lossless compression. This approach is used especially for the efficient lossless on-board compression, data transmission and reconstruction of HSI data. One of the problems for the residual encoding is the speed. In this regard, a recent encoding development showed a throughput which reached 75 GB/s on the GPU. In the works [6, 8, 10] there were resolved two important issues related to the storage and transmission of hyperspectral images, considering the lossless compression models. Also there was studied the SLSQ algorithm and its condition in modern lossless compression schemes, and this algorithm is suitable for on-board implementation. Also two approaches to improve the SLSQ compression are considered. The first approach is a pre-processing of the scheme which orders the channels before compressing a hyperspectral image, also we studied the correlation between the channels. The second approach is based on the pre-processing time and this individualizes a set of channels which are better compressed with two-dimensional inter-channel predictor instead of three-dimensional SLSQ. The main idea of the approach is determined by the fact that the SLSQ NR-groups of the algorithm are not effective, reaching a compression degree of 3,3.

Approaches of lossless compression are considered by most researchers, because the lossless algorithms are effective as the information is important in many fields such as RSE environmental protection, the earth's surface monitoring, military purposes, etc.

Based on the conducted studies and the above approaches to algorithms and methods of compression of hyperspectral AI it can be assumed that these disadvantages stimulate intensive researches aimed at finding new effective solutions which provide high rates of compression degrees and calculators for processing hyperspectral AI, as well as the development (modification) of new algorithmic software.

One of the perspective ways of solving the problems of hyperspectral AI compression is the development of algorithmic software with the new methodology of accounting an inter-channel correlation and creating new algorithms. There are some works devoted to this field [11-13], which suggest the need to develop new algorithms and improve existing ones in lossless compression for hyperspectral AI. There are also the developments in the works [14, 15], where preparatory lossy and lossless processing of algorithms is proposed. In this regard, this article proposes some modifications and new stages of the algorithm of lossless compression based on the conducted researches.

Stages of the compression algorithm

There has been suggested the algorithm with interchannel finding of the best channel of intrabands correlation and the use of indexed lossless encoding, which allows to reduce the dimensions of the image channels and to convert them before compression. As a result, the compression degree of the channels of hyperspectral AI processed by well-known statistical algorithms has become much higher. Also some results of study of the compression degree in the number of channel groups were revealed.

Set of the algorithm stages:

1) the calculation of the correlation matrix between AI channels, finding the best correlated channel pairs in all possible combinations, the subtraction of the best pairs of AI channels;

2) the indexed encoding is the encoding in which the range of hyperspectral data ranges from 0 to 255 bytes, this proposed algorithm reduces the range which varies from 0 to 68. The index is the number of the position of the data range value of additive RGB color model. The effectiveness of indexing and increasing the degree of compression is that the indices are in a small range, so they are well-compressed by entropy encoding;

3) the statistical compression algorithm.

Stages of the algorithm.

Stage 1. Consideration of intrabands correlation.

Calculating the correlation value (intrabands dependence) between hyperspectral AI channels allows to reduce the data range, so fewer number of rates for their storage is needed. The best correlated channel should be found. Then the consideration of intrabands correlation between the selected channels of hyperspectral aerospace image by selecting the correlation value and forming the deviations (differences) of the original data should be taken into account.

For a detailed description of the first stage it is necessary to consider the following objects:

- the original image – matrix of the image values $\mathbf{I}[m,n,k]$, where m,n,k are the indices of lines, columns and channels of the original image, $m=1,2,\dots,M$, $n=1,2,\dots,N$, $k=1,2,\dots,K$;
- $\mathbf{A}[k]$ – the array for values storage when calculating the mathematical expectation of each channel;
- $\mathbf{I}'[m,n,k]$ – the array for placing differences (deviations) values between the value of the mathematical expectation $\mathbf{A}[k]$ and $\mathbf{I}[m,n,k]$;
- $\mathbf{C}[k]$ – the array for storage of correlation values *cor* between the adjacent ranges (channels);
- \mathbf{K}_{best} is the best correlated channel;
- $\mathbf{I}''[m,n,k]$ – the array for placing differences (deviations) values between the channels $\mathbf{I}'[m,n,k]$ and $\mathbf{I}[m,n,k]$;

Step 1. To calculate the mathematical expectation of each channel of the original image $\mathbf{I}[m,n,k]$ and place the values in the array $\mathbf{A}[k]$.

Step 2. To search the channels $\mathbf{I}[m,n,k]$ in all $\mathbf{I}[m,n,k+1]$ and place in the array \mathbf{K}_{best} .

Step 3. To calculate the correlation *cor* (on the basis of $\mathbf{A}[k]$ and $\mathbf{I}[m,n,k]$) for each pair from all available K channels of the original image $\mathbf{I}[m,n,k]$. The result should be placed in $\mathbf{C}[k]$, $k=1,2,\dots,K-1$.

$$\mathbf{C}[k] = \frac{\sum_{m,n} (\mathbf{I}'[m,n,k] - \mathbf{A}[k]) \times (\mathbf{I}'[m,n,k-1] - \mathbf{A}[k-1])}{\sqrt{\sum_{m,n} (\mathbf{I}'[m,n,k] - \mathbf{A}[k])^2 \times \sum_{m,n} (\mathbf{I}'[m,n,k-1] - \mathbf{A}[k-1])^2}}.$$

Step 4. To calculate the differences (on the basis of $\mathbf{A}[k]$, \mathbf{K}_{best} and $\mathbf{I}[m,n,k]$) between the actual values of $\mathbf{I}[m,n,k]$ and the mathematical expectation for each pair from all available K and \mathbf{K}_{best} channels of the original image. The result should be placed in $\mathbf{I}'[m,n,k]$, $k=1,2,\dots,K-1$.

$$\mathbf{I}'[m,n,k] = \sum_{m,n} (\mathbf{I}[m,n,k] - \mathbf{A}[k]).$$

Step 5. To calculate the difference (on the basis of $\mathbf{A}[k]$ and $\mathbf{C}[k]$) between the elements $\mathbf{I}'[m,n,k]$ and the corresponding values of the original image $\mathbf{I}[m,n,k]$ in each from K channels

$$\mathbf{I}'[m, n, k] = \begin{cases} \text{where } C[k] > 0, \mathbf{I}'[m, n, k] - \mathbf{I}[m, n, k - I] \\ \text{where } C[k] \leq 0, \mathbf{I}[m, n, k] \end{cases} \text{ for } m = 1..M, n = 1..N$$

placed in $\mathbf{I}''[m, n, k]$.

As a result the array $\mathbf{I}''[m, n, k]$ based on the intrabands correlation values $C[k]$ is formed.

The subtraction (differential conversion) gives a special advantage in the compression of hyperspectral aerospace images, and also takes into account such spectral characteristics as high interchannel correlation, that increases the compression degree comparing with the universal algorithms which do not consider AI features.

Stage 2. The indexed encoding.

The efficiency of the algorithm is the range of hyperspectral data varied in a certain range of possible values using additional one-dimensional data structures. In data structures the values are kept in the form of indices. Index is the position number of the data range value of hyperspectral AI channel. Then, non-repeating indices of the values of auxiliary data structures are kept instead of the original values of hyperspectral aerospace images.

The formation of the new original data structure by the subtraction of the sequence unit of the values we receive a diminished image thus internal block elements of the image channel are defined, this creates three main data structures.

Let us consider the proposed algorithm with the help of the above stage. We introduce the following objects for consideration:

- auxiliary data structures (ADS) to obtain one-dimensional arrays of each image channel $\mathbf{I}[m, n, k] = \mathbf{I}_1[m, n, k], \mathbf{I}[m, n, k] = \mathbf{I}_2[m, n, k], \mathbf{I}[m, n, k] = \mathbf{I}_3[m, n, k]$, where 1-3 are ADS numbers;
- $\mathbf{I}[k]$ is the array to place the unique indices values of the pixels for each channel $\mathbf{I}[m, n, k]$ and ADS;
- original indices $\mathbf{I}_1[m, n, k], \mathbf{I}_2[m, n, k], \mathbf{I}_3[m, n, k], \mathbf{I}_1'[m, n, k], \mathbf{I}_2'[m, n, k], \mathbf{I}_3'[m, n, k]$.

Step 1. Convert hyperspectral AI channel into ADS $\mathbf{I}[m, n, k] = \mathbf{I}_1[m, n, k], \mathbf{I}[m, n, k] = \mathbf{I}_2[m, n, k], \mathbf{I}[m, n, k] = \mathbf{I}_3[m, n, k]$ corresponding to the values of the original image $\mathbf{I}[m, n, k]$ in the blocks. The result should be put in the arrays $\mathbf{I}'[m, n, k]$.

Step 2. Form the arrays $\mathbf{I}'[m, n, k]$ based on the auxiliary structures by replacing for each $\mathbf{I}_1[m, n, k],$

$$\mathbf{I}_2[m, n, k], \mathbf{I}_3[m, n, k]. \text{ The result is put into the arrays: } \mathbf{I}''[m, n, k]: \mathbf{I}'''[m, n, k] = \begin{cases} \mathbf{I}_1'[m, n, k] \\ \mathbf{I}_2'[m, n, k] \\ \mathbf{I}_3'[m, n, k] \end{cases}.$$

Step 3. Form one-dimensional arrays of the original indices values $\mathbf{I}[k]$ on the basis of arrays

$$\mathbf{I}''[m, n, k] \text{ and ADS. The result is put in } \mathbf{I}[k]: \mathbf{I}[k]: \mathbf{I}[k] = \begin{cases} \mathbf{I}''[m, n, k] \in \mathbf{I}_1 \\ \mathbf{I}''[m, n, k] \in \mathbf{I}_2 \\ \mathbf{I}''[m, n, k] \in \mathbf{I}_3 \end{cases}.$$

In the result, there has been formed the array $\mathbf{I}[k]$ considering the formed array $\mathbf{I}''[m, n, k]$ and intrabands correlation $C[k]$.

The advantage of the proposed algorithm is that the indices are in a small range, due to this they are well-compressed by entropy encoding and standard compression algorithms.

Stage 3. Compression using a well-known arithmetic encoding.

Test hyperspectral data

To determine the effectiveness of the proposed algorithm in terms of compression degree, and also the limits of its applicability there have been conducted the experiments on hyperspectral AI (RSE Aviris). The *AVIRIS* system (*Airborne Visible/Infrared Imaging Spectrometer*) provides simultaneous adoption of the 224 spectral images with a wave length between 400 nm and 2500 nm. Also there has been done the comparison of the proposed algorithm with experimental results obtained for universal archivers (*WinRar*, *WinZip* and *Lossles JPEG 2000* compressor) compression algorithms using an extension of JPEG compression standard, which is widely used in the commercial processing of RSE data. The experiments were done on a PC with *IntelCore i5*, 2.29 GHz and RAM of 4 GB with OS Windows 8.1 (Table).

Table. Characteristics of test hyperspectral AI

Channels number	Image size	Size(byte)	Channels number	Image size	Size(byte)
100	50*50	1081600	200	50*50	2163200
100	100*100	4080400	200	100*100	8160800
100	200*200	16160400	200	200*200	32320800
100	300*300	36240400	200	300*300	72480800
100	400*400	64320400	200	400*400	128640800
100	614*512	125747200	200	614*512	251494400
150	50*50	1622400	224	50*50	2421632
150	100*100	6120600	224	100*100	9140096
150	200*200	24240600	224	200*200	36199296
150	300*300	54360600	224	300*300	81178496
150	400*400	96480600	224	400*400	144077696
150	614*512	188620800	224	614*512	281673728

The results of the experiments

The Fig. 1 shows the compression algorithms when varying the number of channels and the geometric size (50×50) in comparison with the universal archivers *Winrar*, *7Z*, and *JPEG Lossless*. Indicators of compression degree D are 25% higher than universal algorithms through searching the main channel, which determines the priority of channels compression and taking into account interchannel correlation, indexed encoding.

To improve the efficiency of the algorithm, the experiments on the following parameters were done:

- 1) Cor – the given correlation value;
- 2) N – the number of AI channels group;
- 3) R – the size of AI channels;
- 4) K – the number of channels in the group N .

Fig. 2 shows that the indicators of the compression algorithm degree in R are higher in the compression degree than archivers *WinRar*, *7Z*, and *JPEG Lossless* more than 80%.

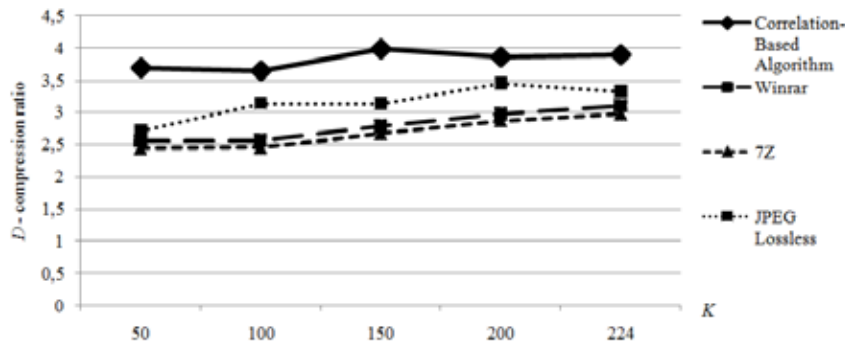


Fig. 1. Lossless compression algorithm according to K

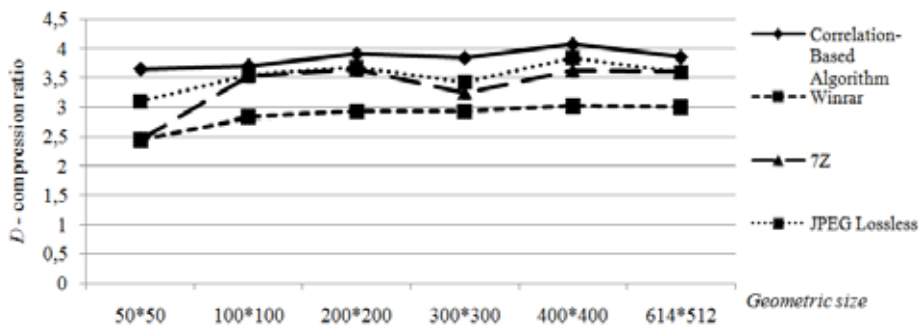


Fig. 2. Lossless compression algorithms at various size

The results of the experiment based on the compression algorithms degree taking into account the correlation and the number of groups show that the recommended parameter values are as follows: $N = 2 - 10$. This demonstrates that the subtraction (differential conversions) are effective with large range of channels in the group, then the average values of differences Sr are the least (Fig. 3), so this allows to store in the smallest amount on a disk space than the small amount of K in N .

In Fig. 4-5 depicts the dependencies on various parameters of the experiments performed showing the modification of the compression ratio, for example, at $N = 5$ ($D > 4$), $N = 9$ ($D > 4,5$).

According to the results of experiments based on the dependence of the correlation from the chosen channels groups and the best channel it is revealed that when the value Cor is reduced then K is reduced proportionally. To study the effectiveness of the proposed compression method with the use of choosing the groups of the most correlated channels, there have been done the experiments on the number of filling channels, which show that when $K = [5:50]$ the compression degree is higher than when $K = [50:100]$, Fig. 6.

Conclusion

On the basis of the conducted researches the following conclusions should be drawn: the compression lossless algorithm considering intrabands correlation allows to increase the compression degree to ($D > 4,5$) than in the universal archivers and JPEG *Lossless* algorithm; the proposed approach of finding the best groups of channels for a given correlation value has increased the efficiency of the

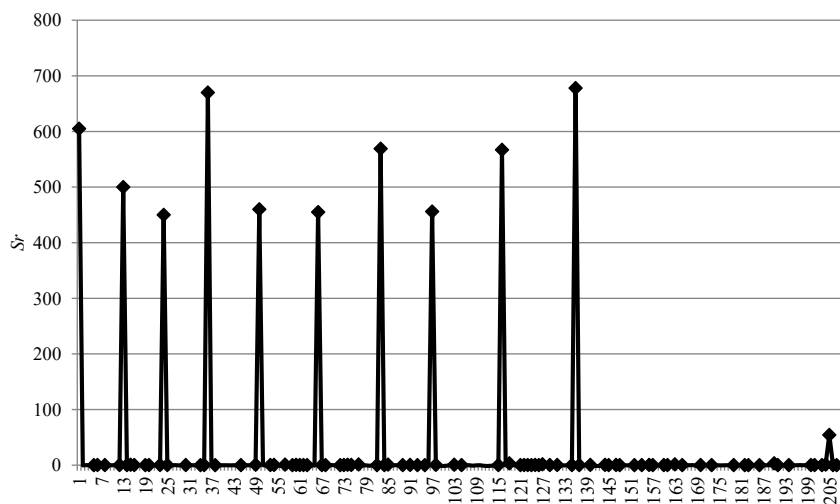


Fig. 3. Sr of the best channels and their differences

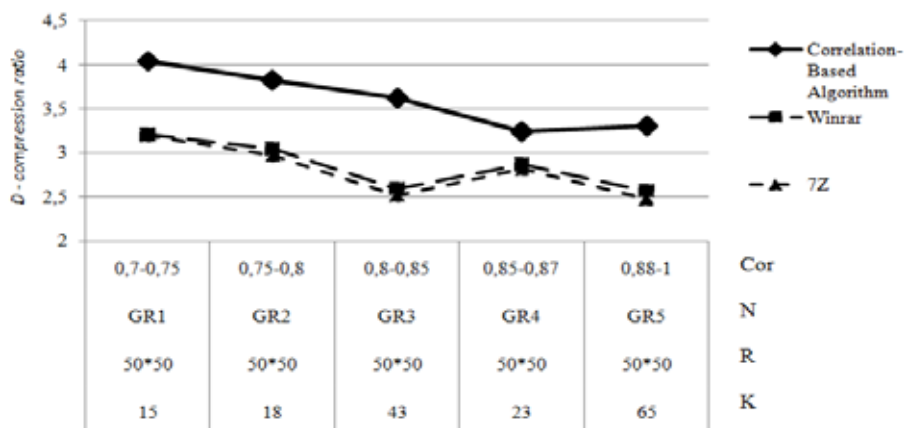


Fig. 4. Comparison of lossless compression algorithms for N = 5

use of channel subtraction phase (differential transformation); the obtained results of the comparison of the converted hyperspectral AI with archivers and JPEG 2000 *Lossless* allow to talk about the efficiency of the indexed encoding stages.

The perspectives for further researches can be encoding the transformed hyperspectral AI with an adaptive algorithm using regression analysis and a new model of interchannel correlation accounting.

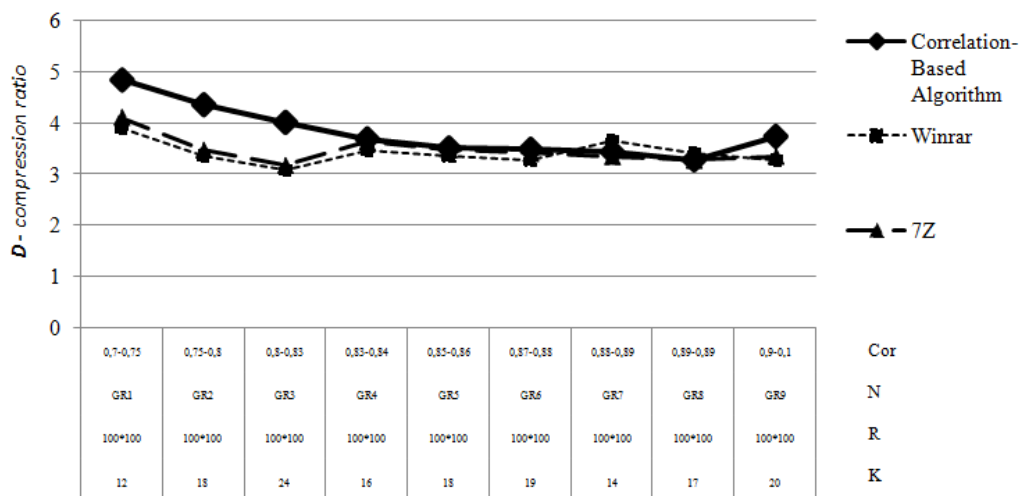


Fig. 5. Comparison of lossless compression algorithms for N = 9

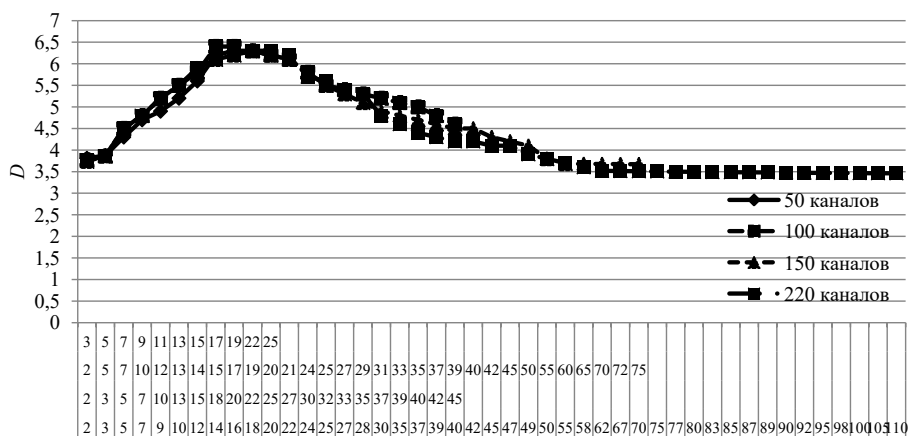


Fig. 6. The graph in number of filling the channels

References

- [1] Yuan L., Jianping L., Ke G. Lossless compression of hyperspectral images using hybrid context prediction. *Optics Express*. 2012, 20(7), 8199-8206.
- [2] Changguo L., Ke G. Lossless Compression of Hyperspectral Images Using Three-Stage Prediction with Adaptive Search Threshold. *International Journal of Signal Processing, Image Processing and Pattern Recognition*. 2014, 7(3), 305-316.
- [3] Cheng-chen L., Yin-Tsung H. Lossless Compression of Hyperspectral Images Using Adaptive Prediction and Backward Search Schemes. *Journal of Information Science and Engineering*. 2011, 27, 419-435.
- [4] Ayman A., Salwa E., Mohamed El. Sh. Hyperspectral Data Compression Using Weighted Spatial-Spectral Lossless Coding Technique. *International Journal of Image Processing*. 2012, 6(6), 468-477.

- [5] Qiang Z., Pauca V.P., Plemmons R. Randomized methods in lossless compression of hyperspectral data. *SPIE Journal of Applied Remote Sensing*. 2013, 7, 1-16.
- [6] Raffaele P., Bruno C. On the Compression of Hyperspectral Data. *IT CoNvergence PRActice*. 2013, 1(4), 24-38.
- [7] Changguo L., Ke G. Lossless Compression of Hyperspectral Images Using Three-Stage Prediction with Adaptive Search Threshold. *International Journal of Signal Processing, Image Processing and Pattern Recognition*. 2014, 7(3), 305-316. <http://dx.doi.org/10.14257/ijsp.2014.7.3.25>
- [8] Ganeshraj P., Sivasankar A. Scalable Compression Method for Hyperspectral Images. *Journal of Engineering Sciences*. 2013, 2, 1-5.
- [9] Diego V., Enrico A. *Novel Rate Control Algorithm for Onboard Predictive Coding of Multispectral and Hyperspectral Images*. 2014. Source: URL: <http://arxiv.org/pdf/1401.3277.pdf>. DOI: 10.1109/TGRS.2013.2296329.
- [10] Raffaele P., Bruno C. Band Clustering for the Lossless Compression of AVIRIS Hyperspectral Images. *ACEEE Int. J. on Signal and Image Processing*, 2014, 5(1), 1-14.
- [11] Algorithm for compressing hyperspectral aerospace images of earth remote sensing / Sarinova A.Zh., Zamyatin A.V. *XI International scientific-practical conference of students, graduate students and young scientists "youth and modern information technologies"*. 2013, 384-386.
- [12] Zamyatin A.V., Sarinova A.Zh. Algorithm for compressing hyperspectral aerospace images with byte processing and interband correlation. *Applied Informatics*. 2013, 5(47), 37-42.
- [13] Sarinova A. Lossless compression of hyperspectral images with pre-byte processing and intra-bands correlation. Sarinova A., Zamyatin A., Cabral P. *Dyna-Colombia*. 2015, 82(190), 166-172.
- [14] Sarinova A. Preparatory processing of algorithms for compressing hyperspectral aerospace images with allowance for band-to-band correlation with losses and without losses. *Proceedings of the XV International Conference named after A.F. Terpugova. Information Technologies and Mathematical Modeling*. 2016, 123-128.
- [15] Zamyatin A.V. Processing the algorithms, compression of hyperspectral images based on aerospace crossband correlation with and without losses. Zamyatin A.V., Sarinova A.Zh. // *Materials of the Nineteenth International Scientific Conference: in 3 volumes. under the Society. Ed. VM Vishnevsky and KE Samuilova Distributed computer and telecommunication networks: control, calculation, communication. (DCCN-2016)*. 2016, 291-298.